Test-Retest Reliability of Reciprocal Isokinetic Knee Extension and Flexion Peak Torque Measurements

Robert W. McCleary, MS, ATC
J.C. Andersen, PhD, ATC, PT

ABSTRACT: Our purpose in conducting this study was to estimate the reliability of reciprocal concentric knee extension and flexion peak torque obtained in uninjured male athletes using the Biodex isokinetic dynamometer. Twenty-six male intercollegiate athletes (age=19.5±4.1 yr; ht=70.3±14.9 in; wt=212.9 ±48.5 lb) participated in this study. We used the Biodex to measure peak torque occurring during right knee extension and flexion over 3 consecutive days. Means and standard deviations were calculated for both extension and flexion on each of the 3 days. We observed a significant main effect for days for both extension and flexion. Mean peak torque for the first day was significantly higher (Tukey, p<.05) than the means for the other 2 days, which were not significantly different from each other (Tukey, p>0.05). Intraclass correlation coefficients (ICCs) were estimated for each of the six trials on each of the 3 days. Intraclass correlation coefficients (ICCs) ranged from .88 on trial 1, day 1 for both extension and flexion peak torque to .97 for extension peak torque and .98 for flexion peak torque on day 3, trial 6. The standard error of measurement for extension peak torque was 7.0 ft-lbs and for flexion peak torque was 3.0 ft-lbs. These results indicate that reliable measurements of reciprocal right knee extension and flexion peak torque can be obtained from uninjured male athletes with the Biodex isokinetic dynamometer.

Robert McCleary is a medical student at the College of Osteopathic Medicine at Ohio University in Athens, OH.

J.C. Andersen is an assistant professor of health science and curriculum director of athletic training education at the College of Health and Professional Studies at Georgia Southern University in Statesboro, GA.

Isokinetic testing is used extensively to measure muscle performance in both research and clinical settings. These measurements are typically made to assess the effect of an experimental treatment or a clinical intervention. For these measurements to be useful, we must have information on the reliability of the measurements when using a specified test protocol in a given subject or patient group (14,16). Therefore, the reliability of these measurements, made using a specific test protocol in the population of interest, is quite important if we plan to use these results in our evaluation of an athlete's functional status (14). In addition, if we plan to make comparisons over time, the measurements taken at any one time should have a high degree of reliability to allow us to determine that a true change in torque output has occurred following some clinical intervention or experimental treatment.

Several recent studies have assessed the test reliability of various protocols using a Biodex isokinetic dynamometer (4,10,13,17). Two of the reliability indices (10, 17), however, were calculated using the Pearson product-moment correlation which is an inappropriate method of assessing test-retest reliability of motor skill or muscle performance measurements (5,11,15). The Pearson product-moment correlation is a measure of association between two different variables and not a measure of agreement between repeated measurements of the same variable (11,15).

More recently, two investigators (4,13) have reported test-retest reliability estimates for peak torque measurements obtained with the Biodex using the more appropriate intraclass correlation coefficient (ICC). Estimates ranged from 0.83 to 0.97 for extension peak torque and from 0.58 to 0.98 for flexion peak torque, over a variety of angular velocities. A mixed sample of males and females participated in both studies and neither study indicated whether their sample included intercollegiate athletes (4,13). These results suggest that, in mixed samples of males and females, reliable measurements of peak torque can be obtained with a specific Biodex testing protocol over a range of angular velocities. Unfortunately, no study to date using the Biodex has indicated whether any of the subjects involved were intercollegiate athletes.

Furthermore, only two studies (4,13) estimated the test-retest reliability of peak torque measurements using the preferred reliability estimation method. We therefore estimated the reliability of reciprocal concentric knee extension and flexion peak torque measurements obtained at 60°/second using the Biodex isokinetic dynamometer in a sample of uninjured male intercollegiate athletes.

Method

Twenty-six male intercollegiate athletes (22 football players and 4 swimmers; age=19.5±4.1 yr; ht=70.3±14.9 in; wt=212.9±48.5 lb) volunteered to participate in this study. Informed consent was obtained.

All subjects were actively involved in their respective team's off-season conditioning program, which included lower quarter weight training. In addition, all subjects had previously undergone a preseason isokinetic screening test and, therefore, had previous experience performing the same movement on the testing device. We screened each subject before participation to ensure that the following criteria were met (6):

1. No history of lower extremity fracture.
2. No history of a neurological condition affecting lower extremity function.

3. No lower extremity muscular strain or ligamentous sprain during the 6 months before testing that limited activity for more than 48 hours.

We scheduled each subject individually for testing at the Athletic Training Room in the Intercollegiate Athletic Department of Ohio University in Athens, Ohio. We used the Biodex B-2000 Clinical Data Station to measure the peak torque occurring during right knee extension and flexion. Before each testing session, we verified calibration of the device according to manufacturer guidelines (1). Gravity correction of the torque measurements was accomplished through the Biodex software package. Knee range of motion (ROM) limits were individually set for each subject with a mean ROM for all subjects of 89.8±2.8°.

During each testing session, we asked each subject to complete a warm-up session of three submaximal contractions followed by three maximal concentric contractions. This warm-up was then followed by 6 maximal test concentric contractions of the knee extensors and flexors for a total of 12 reciprocal contractions at an angular velocity of 60°/second. A 30-second rest period was allowed between the warm-up and test sessions.

Stabilization of each subject was achieved by placing velcro straps across the chest, around the waist, just above the right knee, and just above the right ankle, which secured the right lower leg to the input shaft of the dynamometer. In addition, we visually aligned the estimated transverse rotational axis of the knee with the mechanical axis of the dynamometer. We repeated the entire test procedure for each subject over 3 consecutive days at approximately the same time of day to limit the extent of possible diurnal variation.

We used a two-way, nested random effects model ANOVA (5) to estimate the variance components due to subjects, days, and trials, using the BMDP8V statistical package (3), which then allowed us to calculate the intraclass correlation coefficients. These intraclass coefficients represent ratios of true variance (between subjects) to the sum of the true variance and error variance (variance due to trials and variance due to days). The intraclass coefficient then, is an estimate of the relative reliability for the mean score over any combination of days and trials (15).

We also calculated the standard error of measurement (SEM) to provide an estimate of the absolute reliability of our measurements (15). The SEM is a clinically useful reliability index because it is expressed in the original units of measure. In addition, the SEM can be used to construct a confidence range about an obtained score, which we can use to evaluate whether or not a true change in peak torque has occurred after some clinical intervention.

In this design, only the post hoc analysis of the day's main effect is of statistical interest. This is because we expected that a significant subject main effect would be present and no test of the trial's main effect could be done. Therefore, post hoc analysis of a significant main effect of days, if found, was accomplished using Tukey's HSD for pairwise comparisons (2). An a priori alpha level of 0.05 was used for all tests of significance.

Results

Means and standard deviations for reciprocal knee extension and flexion peak torque measurements obtained at 60°/second over each day are shown in Table 1. There was a significant main effect for days for both extension ($F(52,390)=8.35$, $p<.0005$) and flexion ($F(52,390)=3.71$, $p<.0005$). Mean peak torque on Day 1 was significantly greater than mean peak torque on both Day 2 and Day 3 (Tukey, $p<.05$), while mean peak torque on Day 2 and Day 3 were not significantly different from each other (Tukey, $p>.05$, Table 1).

The percent variance due to days ranged from 4.3% to 7.3% while the percent variance due to trials ranged from 5.8% to 9.5%. As expected, the largest variance component was that due to subjects, which was 87% of the total variation for extension and 86% of the total variation for flexion. Using these variance component estimates, we calculated ICCs of 0.97 for extension peak torque and 0.98 for flexion peak torque, respectively.

Table 2 presents expected ICC values for a given number of days and trials when using the same testing protocol with a similar sample. The SEM for extension peak torque was 7.0 ft-lbs and for flexion peak torque was 3.0 ft-lbs. The 95% confidence range (calculated as 1.96 x SEM) was 13.7 ft-lbs and 5.9 ft-lbs for extension peak torque and flexion peak torque, respectively.

Discussion

Using the Biodex, reliable measurements of reciprocal concentric knee extension and flexion peak torque can be obtained from uninjured male athletes. Our ICC test-retest reliability estimates

Table 1.—Group Means and Standard Deviations for Peak Torque Measurements, in Foot- pounds, over 3 Days at 60°/second (n=26)

	Extension		Flexion	
Day	Mean	SD	Mean	SD
1	162.1	42.6	106.5	24.4
2	152.7	35.9	102.1	22.6
3	154.2	40.3	102.2	22.7

Table 2.—Expected Intraclass Correlations(ICCs) for Six Trials over 3 Days for Knee Extension and Flexion at 60°/second

Day	Condition	Trials					
		1	2	3	4	5	6
1	Extension	.88	.91	.92	.93	.93	.93
	Flexion	.88	.90	.90	.90	.91	.91
2	Extension	.94	.95	.96	.96	.96	.97
	Flexion	.94	.94	.95	.95	.95	.95
3	Extension	.96	.97	.97	.97	.97	.97
	Flexion	.96	.96	.96	.97	.97	.98

are similar to those reported by previous investigators (4,13) for peak torque measurements over a range of angular velocities. Our estimates of the extension and flexion peak torque SEMs are the first to be presented using data obtained from the Biodex; therefore, we are unable to make any conclusions about the absolute reliability of our peak torque measurements in comparison to past investigators (4,13). Our SEM estimates suggest that, to be confident 95 times in 100 isokinetic tests that any observed change in peak torque is real and not due to measurement error or chance events, extension peak torque must change by more than 13.7 ft-lbs and flexion peak torque by more than 5.9 ft-lbs between testing sessions.

Our variance component estimates are consistent with past investigators who used the Biodex. The greatest variance component is due to subjects and reflects true differences in muscle performance among subjects. In our study, the variance due to subjects was less than that reported for extension peak torque, while our results for flexion peak torque are similar to those reported previously (13). However, Montgomery et al (13) measured peak torque at angular velocities of 90°/s and higher while we measured peak torque at 60°/s; thus direct comparisons are not possible. Feiring et al (4) did measure peak torque at 60°/s but did not report variance component estimates.

An unexpected result of our study was the significant main effect of days for both extension and flexion. This result was unexpected because days and trials were assumed to be random sources of error variation and no significant differences over days and trials were expected. Because our subjects were engaged in off-season conditioning that involved lower quarter weight training, it is possible that most, if not all, of our subjects engaged in lower quarter weight training during the testing period. Therefore, the significant reduction in peak torque from Day 1 to Days 2 and 3 could be due to muscular fatigue. Our reliability estimates, however, were not adversely affected (Table 2).

There are several design and methodological issues that affect the clinical significance of our results. We made three critical assumptions concerning our design. We assumed: 1) the variance due to days and the variance due to trials would be random sources of measurement error (5,15); 2) that subjects warmed-up or practiced enough before testing so there would be no trial effect; and 3) there would be no significant "good day-bad day" effect. In view of these assumptions, we observed a significant effect due to days which we believe is associated with our subjects' involvement in off-season weight training. Our finding of a "good day-bad day" effect is an important methodological consideration for clinicians since we did not control for our subjects' involvement in lower-quarter weight training. Controlling for weight training may have eliminated the significant reduction in peak torque from Day 1 to Day 2. However, we felt that the protocol that we used more accurately reflected the reality of our clinical setting where an athlete may have exercised on one day and was tested on the next.

Several authors (8,15) have suggested that the assumption that any variance due to trials in this design is due to random measurement should be reexamined. Although past investigators (8) reported a significant trial effect, more recent studies (7,12) employing this design to assess isokinetic test-retest reliability did not report a significant trial effect. In addition, previous investigators (8) suggested that, to obtain stable peak torque measurements, subjects should warm-up with three submaximal and three maximal contractions. Therefore, since we used the warm-up procedure suggested by other investigators and no trial effect has been demonstrated by past investigators using the Biodex, we believed the randomness of trials assumption was tenable. This assumption may have been inappropriate given our finding of a larger trials variance component than in previous studies.

Several other limitations of our study relate to its clinical relevance. First, we only tested the right knee extensors and flexors of our subjects. Second, since ICC estimates are velocity-dependent (13), we cannot conclude that estimates at angular velocities other than 60°/s will respond as ours did at 60°/s. Finally, perhaps the most important limitation of this study is the small number of subjects sampled, even though our sample (n=26) was the largest sample using intercollegiate athletes employed to date in estimating test-retest reliability of the Biodex. The ICC is a conservative estimator unless the number of subjects is 200 or more (15). Therefore, our ICC reliability estimates are probably smaller than the true population parameters.

Within the limitations discussed above, we conclude that reliable measurements of reciprocal concentric knee extension and flexion peak torque at 60°/s can be obtained in uninjured male intercollegiate athletes using the Biodex isokinetic dynamometer. Furthermore, to obtain reliable measurements of reciprocal peak torque with this protocol, testing should be done over 2 days with between three to six test repetitions on each day. In addition, researchers or clinicians desiring to attain a specified degree of reliability, given a similar sample and identical procedures, could determine an appropriate measurement schedule from Table 2. To be confident that some clinical intervention has resulted in a true change in peak torque, you should see an increase of more than 13.7 ft-lbs in extension peak torque and of more than 5.9 ft-lbs in flexion peak torque. These guidelines are only appropriate, however, when using identical test procedures with similar subjects.

Because of the limitations of our study, further research is needed to determine the influence of a larger sample on the reliability estimates observed using this protocol. The day's main effect must also be evaluated to determine the extent off-season weight training affects reliability estimates. This could be done by repeating our study with a control group not involved in off-season weight training. The assumption of the randomness of trials also should be tested in future studies using the protocol presented here, since our results suggest this assumption may not be appropriate.

References

1. Biodex. System calibration. In: *Clinical Data Station Operations Manual.* Shirley, NY: Biodex Corporation, Inc; 1988.
2. Bruning JL, Kintz BL. *Computational Handbook of Statistics.* 3rd ed. Glenview, Ill: Scott, Foresman, and Company; 1987:124-127.
3. Dixon WJ, ed. *BMDP Statistical Software Manual.* Berkeley, Calif: University of California Press; 1990:1245-1261.
4. Feiring DC, Ellenbecker TS, Derscheid GL. Test-retest reliability of the Biodex isokinetic dynamometer. *J Orthop Sports Phys Ther.* 1990; 11:298-300.
5. Feldt LS, McKee ME. Estimation of the reliability of skill tests. *Res Q.* 1958; 29:279-293.
6. Gross MT, McGrain P, Demilio N, Plyler L. Relationship between multiple predictor variables and normal knee torque production. *Phys Ther.* 1989; 69:54-62.
7. Harding B, Black T, Bruulsema A, Maxwell B, Stratford P. Reliability of a reciprocal test protocol performed on the Kinetic Communicator: an isokinetic test of knee extensor and flexor strength. *J Orthop Sports Phys Ther.* 1988; 10:218-223.
8. Johnson J, Siegel D. Reliability of an isokinetic movement of the knee extensors. *Res Q.* 1978; 49:88-90.

9. Keppel G. *Design and Analysis: A Researcher's Handbook.* 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, Inc; 1982:144-166.
10. Klopfer DA, Greij SD. Examining quadriceps/hamstrings performance at high velocity isokinetics in untrained subjects. *J Orthop Sports Phys Ther.* 1988; 10:18-22.
11. Kroll W. Reliability theory and research decision in selection of a criterion score. *Res Q.* 1967; 38:412-419.
12. Levene JA, Hart BA, Seeds RH, Fuhrman GA. Reliability of reciprocal isokinetic testing of the knee extensors and flexors. *J Orthop Sports Phys Ther.* 1991; 14:121-127.
13. Montgomery LC, Douglass LW, Deuster PA. Reliability of an isokinetic test of muscle strength and endurance. *J Orthop Sports Phys Ther.* 1989; 10:315-322.
14. Rothstein JM. Outcome assessment of therapeutic exercise. In: Basmajian JV, Wolf SL, eds. *Therapeutic Exercise.* 5th ed. Baltimore, Md: Williams & Wilkins; 1990:93-107.
15. Safrit MJ, ed. *Reliability Theory.* Washington, DC: American Alliance for Health, Physical Education and Recreation; 1976:15-35.
16. Task Force on Standards for Measurement in Physical Therapy. Standards for tests and measurements in physical therapy practice. *Phys Ther.* 1991; 71:589-622.
17. Wilk KE, Johnson RD, Levine B. Reliability of the Biodex B-2000 isokinetic dynamometer. *Phys Ther.* 1988; 68:792.